



## Review Article

# Computational methods for metastasis detection in lymph nodes and characterization of the metastasis-free lymph node microarchitecture: A systematic-narrative hybrid review



Elzbieta Budginaite<sup>a,d</sup>, Derek R. Magee<sup>b</sup>, Maximilian Kloft<sup>a,c</sup>, Henry C. Woodruff<sup>d</sup>, Heike I. Grabsch<sup>a,e,\*</sup>

<sup>a</sup> Department of Pathology, GROW - Research Institute for Oncology and Reproduction, Maastricht University Medical Center+, Maastricht, The Netherlands

<sup>b</sup> School of Computing, University of Leeds, Leeds, UK

<sup>c</sup> Department of Internal Medicine, Justus-Liebig-University, Giessen, Germany

<sup>d</sup> Department of Precision Medicine, GROW - Research Institute for Oncology and Reproduction, Maastricht University Medical Center+, Maastricht, The Netherlands

<sup>e</sup> Pathology and Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK

## ARTICLE INFO

## Keywords:

Lymph node  
Artificial intelligence  
Segmentation  
Immunity  
Review

## ABSTRACT

**Background:** Histological examination of tumor draining lymph nodes (LNs) plays a vital role in cancer staging and prognostication. However, as soon as a LN is classed as metastasis-free, no further investigation will be performed and thus, potentially clinically relevant information detectable in tumor-free LNs is currently not captured.

**Objective:** To systematically study and critically assess methods for the analysis of digitized histological LN images described in published research.

**Methods:** A systematic search was conducted in several public databases up to December 2023 using relevant search terms. Studies using brightfield light microscopy images of hematoxylin and eosin or immunohistochemically stained LN tissue sections aiming to detect and/or segment LNs, their compartments or metastatic tumor using artificial intelligence (AI) were included. Dataset, AI methodology, cancer type, and study objective were compared between articles. **Results:** A total of 7201 articles were collected and 73 articles remained for detailed analyses after article screening. Of the remaining articles, 86% aimed at LN metastasis identification, 8% aimed at LN compartment segmentation, and remaining focused on LN contouring. Furthermore, 78% of articles used patch classification and 22% used pixel segmentation models for analyses. Five out of six studies (83%) of metastasis-free LNs were performed on publicly unavailable datasets, making quantitative article comparison impossible.

**Conclusions:** Multi-scale models mimicking multiple microscopy zooms show promise for computational LN analysis. Large-scale datasets are needed to establish the clinical relevance of analyzing metastasis-free LN in detail. Further research is needed to identify clinically interpretable metrics for LN compartment characterization.

## Contents

Introduction . . . . .	2
Methods . . . . .	3
Search strategy . . . . .	3
Selection criteria . . . . .	3
Screening by title and abstract . . . . .	3
Full text screening . . . . .	3
Study quality assessment. . . . .	4
Results . . . . .	4
Datasets . . . . .	4
Lymph node contour detection . . . . .	4
Lymph node metastasis detection . . . . .	6
Studies using convolutional neural network models . . . . .	6
Studies using convolutional autoencoders . . . . .	6

\* Corresponding author.

E-mail address: [h.grabsch@maastrichtuniversity.nl](mailto:h.grabsch@maastrichtuniversity.nl) (H.I. Grabsch).

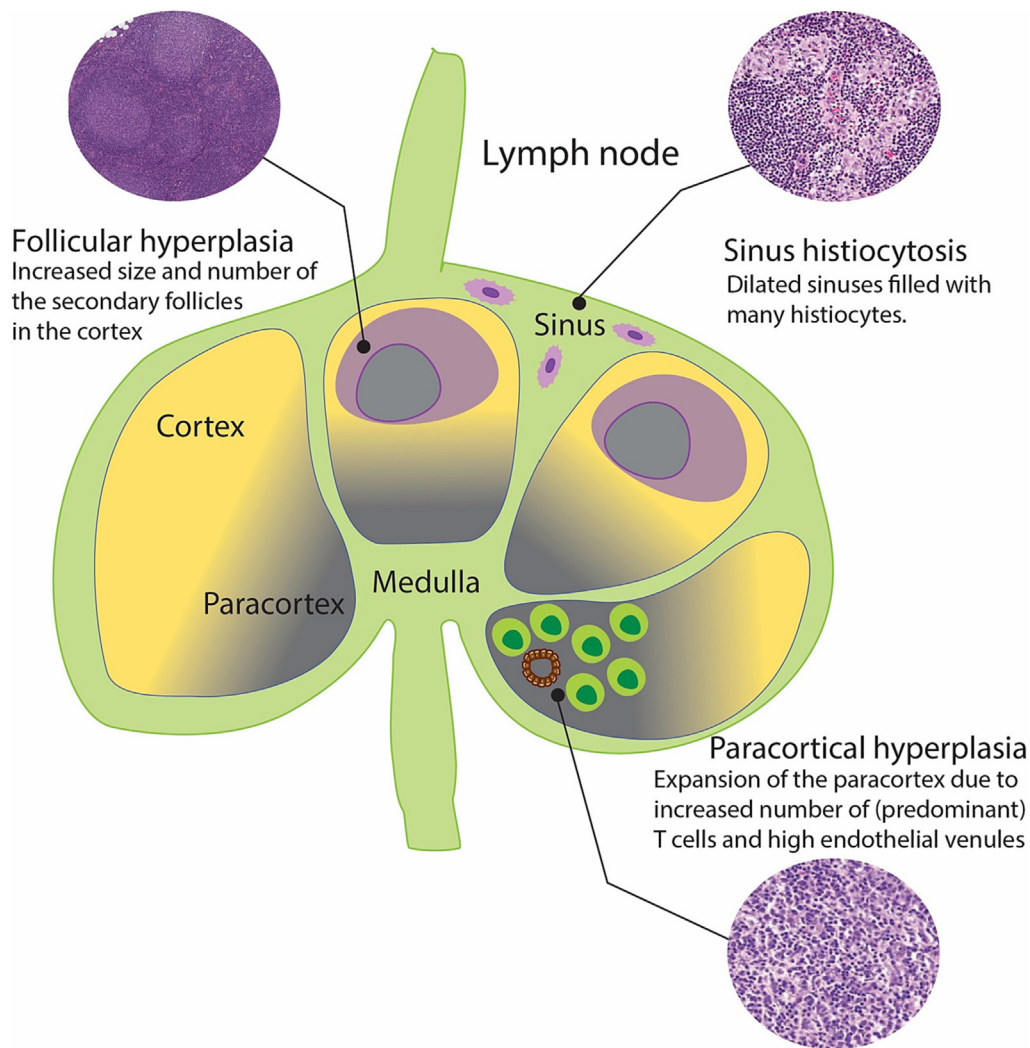
Multi-scale models . . . . .	7
Generative adversarial networks . . . . .	7
Multiple instance learning . . . . .	7
Other machine learning methodologies . . . . .	7
Interpretation of LN metastasis misclassification . . . . .	8
Segmentation of lymph node compartments . . . . .	8
Model selection and experimental design. . . . .	8
Chronological comparison of data and AI strategy selection. . . . .	8
Discussion . . . . .	8
Limitations. . . . .	10
Conclusions . . . . .	10
Declaration of competing interest . . . . .	10
Acknowledgments . . . . .	10
Appendix A. Supplementary data . . . . .	10
References . . . . .	10

**Introduction**

The histological examination of lymph nodes (LNs) is an essential part of cancer staging. The LN status (presence or absence of metastasis) is considered a key prognostic factor in patients with various different cancer types, including breast, gastric, esophageal, and colorectal.<sup>1-4</sup> The number of regional LNs with metastases as well as the extent (size) of the metastatic

lesion within each LN, which may vary from a macroscopically visible lesion to a micro-lesion consisting of a few isolated tumor cells, are the main characteristics defining the regional spread of the cancer.<sup>5</sup>

There is evidence accumulating which suggests that metastasis-free LN microarchitecture rearrangements might be clinically important.<sup>6-8</sup> The LN microarchitecture can be divided into distinct layers exhibiting different compositions and functions, see Fig. 1. The outer layer, named cortex,



**Fig. 1.** Schematic representation of immunologically stimulated lymph node with enlisted compartment-specific reactive patterns.

consists of primary follicles composed of naïve B-cells. Upon immune stimulation, these primary follicles can transform into secondary follicles with germinal centers where further B-cells proliferate. The next (deeper) LN layer, named paracortex, typically contains T-cells and dendritic cells which enter the LN via high endothelial venules (HEVs). The innermost layer of LN, called medulla, containing plasma cells, B-cells and macrophages, opens into hilum and efferent lymphatic vessels. The LN has a conduit system (so called sinuses) that transports macrophages and antigens from outside the LN to the inside. Changes in the LN can manifest as so-called reactive patterns which are illustrated in Fig. 1. The presence of structural changes such as sinus histiocytosis has been associated with the presence of a host anti-tumor response in patients with breast, gastric, and colorectal cancers.<sup>9–11</sup> In the same cancers, the presence of LNs with follicular hyperplasia were related to a longer patient survival.<sup>12–14</sup> It has been suggested that the presence of a hyperplastic LN paracortex might be a predictor of better survival in gastric cancer patients.<sup>15</sup> It has also been suggested that HEVs, usually present in the LN paracortex, may undergo morphological remodeling in pre-metastatic LNs and hence could be a potential morphological biomarker for prediction and prognosis in cancer patients.<sup>16</sup>

To validate the clinical importance of reactive LN patterns or reactive LNs as a whole, large scale investigations need to be performed on sufficiently large cohorts. In order to obtain the measurements of individual LN compartments, these would need to be segmented manually which is very time consuming. Thus, in order to generate large-scale evidence, there is a need for automated LN compartment analysis methods. As LN reaction patterns can be seen through a microscope (see Fig. 2), digital histological images containing LNs should be suitable for artificial intelligence (AI)-assisted analysis.

The launch of the Camelyon16 challenge in 2016 marked the start of a very productive period for researchers investigating computational methods for metastatic LN tumor characterization. The Camelyon datasets consist of 1399 digitized axillary LN slides with partial metastatic tumor annotations and patient-level pathological lymph node (pN) status.<sup>17</sup> The slides from Camelyon datasets were collected from multiple centers and thus vary in stain intensities and sample preparation protocols, reflecting the practical challenges facing AI models such as generalizing across multi-center data. Since the announcement of the Camelyon challenges,<sup>17</sup> there has been a relatively large number of publications proposing different algorithms to detect LN metastases using the publicly available whole slide images (WSI) of axillary LNs from breast cancer patients. However, the computational characterization of metastasis-free LNs seems to be lagging

behind, perhaps due to the fact that there are no well-established benchmarks like the Camelyon challenge and no publicly available annotated datasets. This review aims to systematically identify and analyze recent publications that applied some form of AI for LN characterization, focusing on the identification of AI methodologies in LN analysis, clinical interpretability of the AI-based results, and current shortcomings of AI tools potentially impeding clinical implementation.

## Methods

### Search strategy

This review was conducted following the proposed reporting items for systematic reviews and meta-analyses (PRISMA) protocol.<sup>18</sup> We searched for publications across the following databases: PubMed, Scopus, ArXiv, Web of Science, and Institute of Electrical and Electronics Engineers Xplore, and the search phrases are listed in the supplementary document S1. The databases were searched for articles published up to December 1, 2023.

### Selection criteria

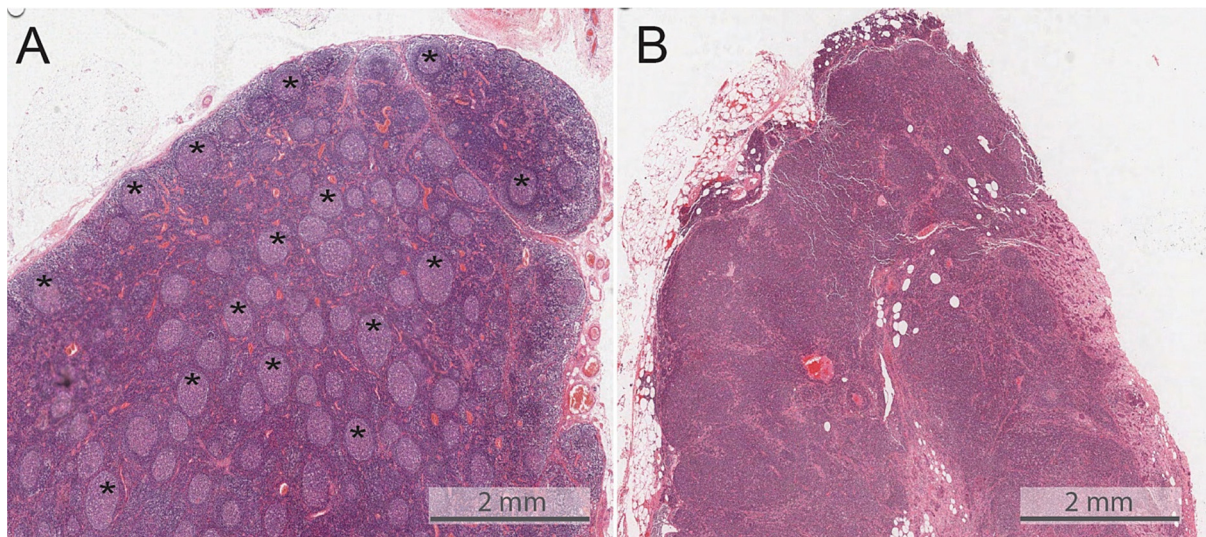
Studies focusing on lymph node (LN) detection, computational LN characterization, or compartment segmentation in histological whole slide images (WSI), including hematoxylin and eosin (H&E) and immunohistochemistry (IHC) staining using AI methods, either deep learning or machine learning, were included in our analysis. Studies reporting on imaging modalities other than histopathology, such as positron emission tomography, computer tomography, ultrasound scans, tissue simulation studies, and tumor immune microenvironment studies were outside the scope of this review and were thus excluded, as well as studies published in other languages than English.

### Screening by title and abstract

To select relevant studies eligible for full text screening, the initially identified studies were filtered based on title and abstract content considering the above-mentioned inclusion/exclusion criteria.

### Full text screening

The full text of eligible articles was further analyzed to identify and exclude out of scope articles. Articles with unretrievable full text, as well as



**Fig. 2.** Exemplary images taken from H&E stained lymph node from esophageal cancer patients. The figure illustrates the visually perceivable differences between immunologically stimulated (also called reactive) LNs. Figure A shows a lymph node with a large number of germinal centers (some are highlighted in asterisk), indicative of follicular hyperplasia; Figure B contains a metastatic lymph node with hyperplasia of paracortex and a lack of germinal centers.

conference abstracts were excluded from further processing. Data such as dataset, AI model, cancer type, and study objective were extracted from each full text for further article assessment.

### Study quality assessment

In order to assess the experimental design of the study and risk of bias, quality assessment was performed following the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) protocol.<sup>19</sup> For quality assessment, a detailed description of the AI model used in the study as well as clearly defined training data were deemed mandatory for the study to be included in the current review.

## Results

Following the inclusion and exclusion criteria, the PRISMA protocol for systematic reviews, and the CLAIM protocol for study assessment, we finally identified 73 studies to be included in our systematic review, see Fig 3 for PRISMA flowchart and Supplementary Table S2 for CLAIM assessment results.

### Datasets

56% ( $n=41$ ) of the studies included in this review utilized publicly available datasets (see Supplementary table S3). 36% ( $n=26$ ) of the studies used the Camelyon16 dataset,<sup>20</sup> consisting of 399 H&E stained breast cancer sentinel lymph node (SLN) WSIs with LN status labels (with metastases versus without metastases) and partial pixel annotations of metastatic

lesions. The second most widely utilized dataset was Patch Camelyon,<sup>21</sup> which was used by 8 (11%) studies and which is a derivative patch version of Camelyon16 (see Table 1). Eight (11%) studies used Camelyon17,<sup>17</sup> a multi-center dataset extending Camelyon16 to 1399 breast cancer SLN WSI. Two (3%) studies utilized the SLN breast cancer dataset from the cancer imaging archive (TCIA) with slide based LN status labels.<sup>22</sup> One study reported results obtained from an axillary LN breast cancer dataset available from the Analytic Image Diagnostics Arena (AIDA) Data Hub, which provides slides immunohistochemically stained for cytokeratin (clone AE1/AE3) instead of manual annotations as ground truth for cancer metastasis.<sup>23</sup> There were two publicly available LN datasets that did not focus on metastasis detection. The dataset published by Bekkhus et al<sup>16</sup> was used for HEV segmentation in immunofluorescently stained LN slides from breast cancer including metastatic and metastasis-free LNs. Furthermore, Gamez Serna et al published a dataset with their multi-magnification organ network (MMO-Net) that included rat mandibular LNs with annotated LN contours.<sup>24</sup> More details on each dataset are provided in Table 1.

32 studies (44%) utilized publicly unavailable in-house datasets. The use of in-house datasets increased steadily overtime as can be seen in Fig. 4C.

### Lymph node contour detection

From an anatomical perspective, LN contour detection is the most general task in computational LN analyses. We identified eight studies proposing automated LN detection solutions. An early adoption of machine learning techniques was used in a study by Niemisto et al.<sup>25</sup> In 2005,

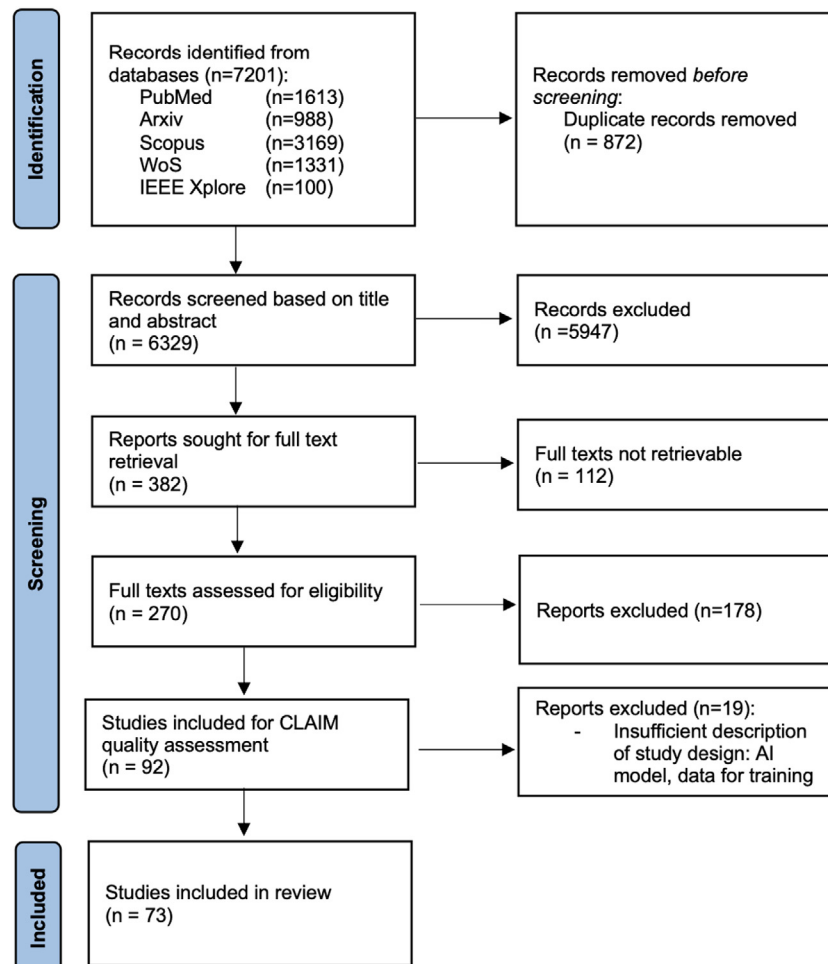


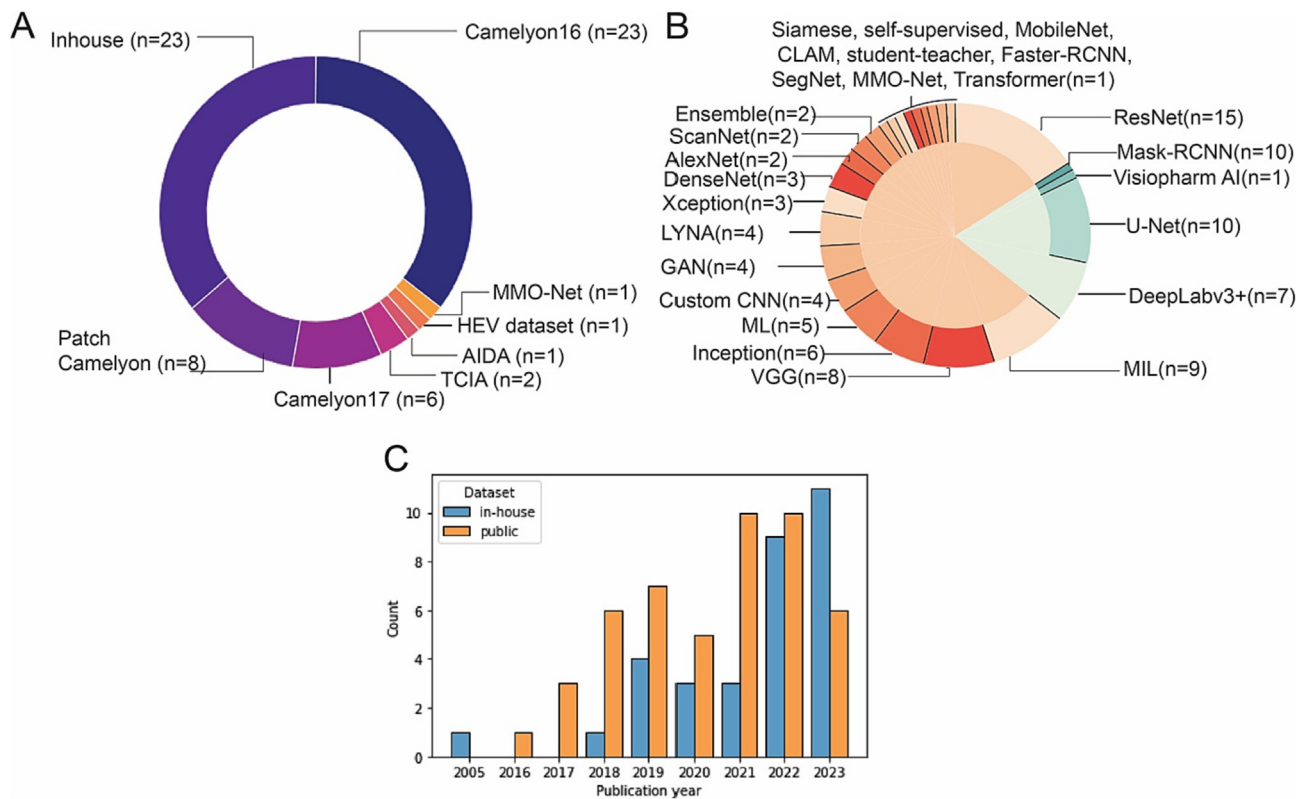
Fig. 3. PRISMA flowchart illustrating the article screening and selection process implemented in this systematic search.

**Table 1**  
Summary of publicly available datasets containing LN tissue sections.

Dataset	Summary	Ground truth	Sample size	Dimensionality	Suitable applications
Camelyon16	H&E stained slides of breast cancer sentinel lymph nodes from 2 centers	Patient-level labels: pN stage. Pixel-level: partial tumor annotations	399 WSIs	WSIs scanned at 20x and 40x magnifications	Segmentation, slide-level classification, MIL
Camelyon17	H&E stained slides of breast cancer sentinel lymph nodes from 5 centers	Patient-level pN stage label. Pixel-level: partial tumor annotations	1000 WSIs	WSIs scanned at 20x and 40x	Segmentation, classification, MIL
Patch Camelyon	Patches extracted from Camelyon16 dataset	Patch-level label: tumor positive or negative	327,680 patches	96x 95 pixel patches at 10x	Patch classification
SLN-Breast, TCIA	H&E stained slides scanned in one center	Patient-level labels: pN stage label	130 slides from 78 patients	WSIs scanned at 20x	Slide-level classification, MIL
Axillary lymph nodes in breast cancer, AIDA	H&E slides of breast cancer sentinel lymph nodes (2 consecutive slices) from multiple scanners	Patient-level labels: treatment, pN stage, primary tumor. IHC slides for cytokeratin AE1/AE3	396 patients	WSIs scanned at 20x	Tumor segmentation, slide-level classification, MIL
Breast cancer lymph nodes for HEV detection	Immunofluorescently stained HEVs using HEV markers for peripheral node addressin PNAd and vascular marker Claudin-5	Immunofluorescence signal for HEVs	73 patients with invasive, non-invasive breast carcinoma, or cancer-free	Not specified	HEV segmentation and diameter evaluation in LNs
Multi-magnification organ network (MMO-Net) dataset	H&E stained slides of rat mandibular lymph nodes from 3 centers	Manual LN contour annotations	53 WSIs containing lymph nodes, 267 WSIs from other organs	Not specified	LN contour segmentation

where K-means clustering algorithm was used to segment LNs based on a simplified three-color clustering scheme. Verghese et al<sup>26</sup> applied a modified Otsu thresholding method to localize LNs in the whole slide breast cancer images. In the study by Wang et al,<sup>27</sup> a U-Net autoencoder was trained using 1x magnification regional gastric cancer LN images to localize the LNs within the slide. A more sophisticated approach for LN segmentation was recently reported by Gamez Serna et al,<sup>24</sup> where a DenseNet121-derived MMO-Net model was trained to detect rat mandibular LNs in a multi-scale fashion accounting for two different image

magnifications in parallel (5 and 1.25). Huang et al<sup>28</sup> trained DeepLabv3+ segmentation model for LN detection in gastric cancer patients at 1.25x magnification, whereas Hu et al<sup>29</sup> utilized a pipeline of faster recurrent CNN (Faster-RCNN) and DeepLabv3+ models for LN bounding box detection followed by precise LN contour delineation. A two-step LN detection approach has also been implemented by Beuque et al,<sup>30</sup> where a U-Net model was trained to obtain LN mask from WSI thumbnail, followed by false-positive prediction filtering by XGBoost model trained on hand-crafted radiomics features obtained from LN masks. Faster-RCNN model



**Fig. 4.** Graphs illustrating the usage patterns of the datasets and models in selected studies. A: piechart with dataset distribution across the selected papers, B: piechart illustrating the model selection in the included articles. The total number of models is larger than the number of studies, as some studies trained more than one model. C: chronological chart of dataset usage focusing on inhouse/public dataset selection. The barchart illustrates the steady increase of studies using inhouse datasets.

was also used for LN detection by Tan et al,<sup>31</sup> where a model was trained to detect LN bounding boxes in colorectal cancer patients at 5x magnification.

### *Lymph node metastasis detection*

Metastatic LN status identification is essential for predicting cancer patient's prognosis and determine further treatment. In our systematic review, 63 (86%) out of 73 studies focused on metastatic lesion detection in regional LNs. 48 out of those 62 studies reported results from analyzing metastatic breast cancer regional LNs. This bias towards LN metastasis detection in breast cancer can be largely explained by the availability of public datasets, namely Camelyon16 and Camelyon17. Based on the machine learning paradigm utilized in a study, we grouped the metastasis detection articles into six groups.

#### *Studies using convolutional neural network models*

The use of convolutional neural networks (CNNs) has been widely adopted in AI-assisted image analysis field. The CNN architecture relies on convolutional filters called kernels—small matrices with learnable weights. The convolution process entails a sequential kernel application across the input image matrix, where a feature map is created via element-wise multiplication of the kernel and a subset of input matrix.<sup>32</sup> There were multiple CNN architectures suggested throughout the years which have been universally applied in computer vision tasks. In 2014, Google has introduced their GoogleNet model, also called Inception<sup>33</sup>—a CNN model that suggested a parallel utilization of multiple kernels of different size. A year after ResNet50 architecture was suggested,<sup>34</sup> solving the vanishing gradient issue via skip (residual) connections. The original model contained 50 layers, with alternative ResNet101 (101 layers) and ResNet152 (152 layers) versions. Vanishing gradient problem was also addressed by DenseNet model,<sup>35</sup> where each layer receives an input from all prior layers via dense connections.

Out of 63 articles discussing metastatic LN analysis, 27 papers were dedicated to a classic deep learning strategy for computer vision, where one or multiple CNN architectures were trained and tested on LN metastasis detection datasets. The earliest studies included in our review starting in 2016 applied commonly used CNN architectures such as AlexNet, DenseNet, GoogleNet and Visual Geometry Group CNN (VGG), as well as custom architectures like ScanNet on Camelyon 16 dataset.<sup>36–41</sup> The Camelyon16 dataset was further utilized to extract metastatic and metastasis-free LN patches into a dataset called Patch Camelyon, which was applied for patch classifier models.<sup>21</sup> Later Camelyon dataset usecases were adapted to address practical WSI processing challenges. WSI scanning speed was addressed by Zhang et al,<sup>40</sup> where authors proposed a multiple spatial context network and patch feature sharing for faster slide scanning. Shvetsova et al<sup>42</sup> proposed unsupervised anomaly detection in Camelyon16 using an autoencoder reconstruction of the input image focusing on the evaluation of the perceptual loss. Alheejawi et al<sup>43</sup> proposed a deep learning method for proliferative index (PI) evaluation in LNs of melanoma cancer in Ki-67 stained WSIs. The authors used a SegNet model to obtain pixel-level segmentation masks. Hu et al<sup>29</sup> utilized inhouse gastric cancer dataset to train DenseNet121 model for metastatic LN detection. Pham et al<sup>44</sup> trained a pipeline of two VGG models for germinal center and metastasis detection, concluding that additional secondary follicle detection step improved tumor model specificity. Allam et al<sup>45</sup> suggested a novel approach towards LN metastasis identification in breast cancer patients—instead of predicting metastasis directly from metastatic regions, authors trained a custom architecture CNN model to detect metastasis from non-metastatic LN regions (metastatic tumor microenvironment). The authors evaluated LN sinuses, follicles, and interfollicular lymphocytic areas concluding that lymphocyte-rich interfollicular regions had the highest positive predictive power for metastatic LN detection.

13 studies (18%) utilized ResNet model architecture to detect LN metastasis. Out of these studies, five articles utilized Camelyon datasets. Jaiswal et al<sup>46</sup> trained classical patch classifiers such as ResNet, VGG, Inception on

Patch Camelyon in semi-supervised fashion using class with maximum predicted probability as ground truth labels for unlabeled patches. The authors used test time augmentation technique to expand the test dataset and evaluate model generalizability. Kim et al<sup>47</sup> used the Camelyon datasets on the Inceptionv3, VGG16, and ResNet models as transfer learning for their in-house dataset of frozen H&E breast cancer LN samples validating that the features learned on Camelyon datasets were transferable to frozen tissue samples. In a study by Chen et al,<sup>48</sup> the authors used a RRCART model to distinguish high-accuracy metastasis predictions from low-accuracy predictions obtained using ResNet model. Lee and Paeng et al<sup>49</sup> passed a tumor probability map generated with ResNet101 into a random forest model for pN prediction, achieving second best-weighted kappa score of 0.9203 within the pool of selected studies leveraging Camelyon17 dataset. Patil et al<sup>50</sup> introduced HistoROI model—a derivative of ResNet architecture that was designed to keep the diagnostician in the loop via active learning. Turki et al<sup>51</sup> trained multiple models, namely ResNet, DenseNet121, VGG16, and Xception on the SLN-breast dataset from TCIA, concluding that DenseNet121 achieved the highest performance for their dataset of choice.

Wang et al<sup>27</sup> used the ResNet50 binary patch classifier for gastric cancer regional LN segmentation, using germinal center, sinuses, and adipose tissue annotations for the identification of the metastasis-free LN class. This study also analyzed the spatial distribution of metastatic lesions, specifying two distinct metastatic dissemination patterns in gastric cancer patients. Chuang et al<sup>52</sup> utilized ResNet50 for metastasis identification in colorectal LNs from colorectal cancer patients coupled with class activation mapping to highlight diagnostically important regions for the model. ResNet model has been utilized for LN metastasis detection in gastric cancer patients by Huang et al<sup>28</sup> and Matsushima et al<sup>53</sup> who concluded that utilization of ResNet model significantly increased the micrometastasis detection sensitivity compared to manual unassisted analysis, however, the automated slide analysis took longer than time needed by pathologist. In the study by Kronberg et al,<sup>54</sup> authors trained a multi-class model for metastatic LN detection in pancreatic cancer patients, capable of distinguishing between healthy pancreas, metastasis-free LN, metastasis, background, and adipose classes.

When designing an automated diagnostic tool, it is important to minimize the chance of false-negative decision, where a metastatic LN is assigned to a “healthy” metastasis-free class, which is reflected by sensitivity metric. Two studies in our analysis proposed ResNet-based solutions that achieved a perfect 1.00 sensitivity in gastric cancer<sup>55</sup> and head and neck cancer<sup>56</sup> datasets consisting of 40 gastric cancer patients and 50 head and neck cancer images, respectively. The proposed metastasis detection systems reached lower yet still competitive false-positivity (specificity) scores of 0.9994 in gastric cancer and 0.759 in head and neck cancer studies, respectively.

#### *Studies using convolutional autoencoders*

Autoencoders are a specific type of CNN networks consisting of encoder and a decoder part aiming to learn a condensed representation of input image. The encoder part compresses the input image into an embedding vector in the autoencoder bottleneck, whereas a decoder part is responsible for input image reconstruction from the embedding latent space. In contrast to the previously described CNN classifier models, which assign a class to the input image patch, autoencoders can produce a finer output mask by classifying each pixel within an input image into a predefined class. In 2015, Ronnenberger et al<sup>57</sup> introduced a U-Net—a U-shaped autoencoder architecture containing skip connections between encoder and decoder parts to help preserve the spatial information. Two years later Google introduced a DeepLabv3+ model.<sup>58</sup> The DeepLabv3+ architecture utilizes patch classifier models such as ResNet50 or DenseNet for feature extraction followed by dilated (also called atrous) convolutions for a wider receptive field and contains an atrous spatial pyramidal pooling (ASPP) module designed for multi-scale feature extraction via multiple atrous convolution rates.

Eight studies utilized pixel-level segmentation models, namely U-Net, DeepLabv3+, or their derivatives. Xu et al<sup>59</sup> leveraged a U-Net model to segment metastases in LNs of the Camelyon16 and Camelyon17 datasets followed by random forest model for pN stage prediction. Jin et al<sup>60</sup> introduced ConcatNet model by concatenating four U-Net models each for nucleus, mitosis, epithelium, and tubule segmentation. Jansen et al<sup>61</sup> utilized U-Net model with ResNet50 encoder for metastasis identification in LNs from melanoma patients achieving competitive sensitivity scores of 91.67 and 95.62 on two datasets. Mainovskaya et al<sup>62</sup> used a pixel segmentation approach using DeepLabv3 model for metastasis segmentation in CRC LNs. The authors also considered the presence of slide artifacts such as tissue folds and pigment residues, which were addressed by training an additional U-Net model for artifact detection. DeepLabv3+ model was applied for metastasis identification in breast<sup>63</sup> and esophageal<sup>64</sup> cancer studies. Bozdag et al<sup>65</sup> addressed the large parameter space of DeepLabv3+ model and introduced a lightweight pyramid-structured segmentation network NonLocalSeg for more time-efficient slide processing in the Camelyon16 dataset. Wang et al<sup>66</sup> utilized a modified DeepLabv3+ model by introducing pyramidal attentional atrous spatial pyramidal pooling (PA-ASPP) and scale-aware selection to weight multi-scale features extracted by ASPP using ResNet101 model as feature extractor. To obtain slide-level metastasis predictions, the authors used density based spatial clustering model (DBSCAN) to combine patch-level predictions and XGBoost model for patient-level pN prediction. The proposed model framework achieved weighted kappa score of 0.9632 which was the highest score among included articles utilizing Camelyon17 dataset in this review.

#### Multi-scale models

Among the selected study pool, we identified five articles that used multi-scale models. Multi-scale models, originally proposed by Wang et al in 2016,<sup>67</sup> leverage tissue context information by ingesting two input images of different magnification at the same time, thus mimicking multiple microscopic zoom levels. In 2017, Liu et al presented Google's Lymph Node Assistant (LYNA) model trained on Camelyon16 patches.<sup>68,69</sup> The LYNA model was built on Inception V3 architecture in multi-scale fashion. Steiner et al<sup>70</sup> reported that the use of the LYNA model as digital assistant for pathologists significantly reduced the slide screening time needed to reach a final diagnosis. The authors have also tested the LYNA framework on their in-house dataset and reported the LYNA model robustness against slide artifacts such as bubbles and H&E stain variability. In a study by Schmitz et al,<sup>71</sup> the authors investigated multi-scale WSI training strategies, achieving higher weighted Jaccard coefficient values for the Camelyon16 dataset compared to the use of a single-scale U-Net model. In recently published study by Wang et al,<sup>72</sup> authors have also utilized Camelyon16 dataset for their multi-scale model, that was based on ResNet architecture. The authors proposed a model consisting of low- and high magnification networks operating at 10x and 40x magnifications, respectively. The model was designed to first scan the slide at low magnification to identify regions of interest that would further be analyzed at high magnification. The proposed multi-scale model was compared to LYNA model with respect to the testing FROC metric and processing time, achieving higher FROC performance with significantly lower slide processing time (33 min per slide for LYNA model compared to 4.5 min for proposed model).

#### Generative adversarial networks

To expand a particular given dataset and improve deep learning model generalizability, image augmentation is often employed, including image rotation, axis flipping, and/or color saturation variations. A more advanced method to augment the dataset is by leveraging generative adversarial networks (GANs)—a specific family of deep learning models trained to generate images from random noise—for synthetic data generation. Four studies trained GANs on the Camelyon datasets. Kovalev et al<sup>73</sup> employed two GAN models, namely deep convolutional GAN and progressive growing GAN to extract deep features from the synthetic images and trained classic machine

learning models for image classification into either metastatic or metastasis-free image patch. Jiarong et al<sup>74</sup> introduced HistoGAN model for synthetic image generation which was further applied in a study by Xue et al<sup>74</sup> proving that synthetic image generation improves classification model performance. Due to the usually site-specific sample preparation protocols, staining intensity variability, tissue thickness variability, and scanner-specific artifacts, it is challenging to build a generalized model capable of performing the task equally well using slides from different centers. Wollmann et al<sup>75</sup> explored this issue by using a CycleGAN model for the domain adaptation task. The generative model was trained to transfer the medical center-specific styles to the slides from new medical centers unseen by the deep learning model aiming to reduce inter-center image variability.

#### Multiple instance learning

The development of deep learning systems usually requires large amounts of data with labeled ground truth, which is a challenge in pathology field due to the limited time resources of pathologists that annotate the data and large WSI scale. The large WSI dimensions (on average, 100.000 x 100.000 pixels)<sup>76</sup> complicate the annotation process, rendering the full WSI manual annotations a barely realistic option. Camelyon datasets are a good example of annotation sparsity, where only partial pixel annotations are available accompanied with a patient-level label. The challenges of sparse data annotations can be facilitated by applying multiple instance learning (MIL) paradigm. MIL training is a branch of weakly supervised model training approaches, where the training instances are bagged in batches with a single label per bag. For our analysis, we identified nine articles implementing this approach, eight of which leveraged Camelyon datasets. Akbar et al<sup>77</sup> implemented a MIL model for Camelyon16 dataset by training a variational autoencoder in unsupervised manner and subsequently clustering the learned features into pre-defined labels. A categorical cross-entropy loss modification was introduced for weak label assignment. Courtiol et al<sup>41</sup> implemented MIL via learning the top instances and negative evidence. Shao et al<sup>78</sup> utilized visual transformer model for transMIL approach. The same authors produced attention heatmaps to visualize the image areas deemed important by the AI algorithm, rendering the result more explainable. Wang et al introduced a second-order MIL relying on the matrix power normalized covariance pooling, so-called second-order feature extraction, from LN slides. The authors applied a channel attention mechanism to identify the most discriminative second order features in the slide. Wang et al<sup>79</sup> introduced a label cleaning MIL approach for sparse annotation refinement using only single WSI for model training, thus suggesting an alternative to the regular data-intensive deep learning training protocols. Kang et al<sup>80</sup> used transfer learning technique for their MIL approach by training the model jointly on Camelyon dataset and their in-house esophageal cancer dataset. The use of publicly available dataset reduced the need for high-quality inhouse annotations. Sadafi et al<sup>81</sup> implemented attention-based MIL approach with active learning. The proposed system evaluated the uncertainty of classification results submitting the slides with the least certain predictions to the expert for additional annotations. Yu et al<sup>82</sup> proposed vocabulary-based MIL approach, where the model was trained to discover the structural prototypes metastatic LNs from breast cancer patients via unsupervised clustering, thus offering a higher model interpretability. Tan et al<sup>31</sup> proposed a transformer-based MIL model for colorectal cancer cases. The authors emphasized that the attention map obtained with the transformer model correctly localized the metastatic LN lesions even if the final prediction was negative, suggesting the model's capability of avoiding false negatives. Transformer model has also been utilized by Qin et al,<sup>83</sup> achieving a higher sensitivity in Camelyon16 dataset compared to pathologist (94.25% and 73.2%, respectively).

#### Other machine learning methodologies

In the field of computer assisted diagnosis, error rate minimization is of paramount importance. One way to reduce the variance of predicted model

errors is to use a combination of multiple machine learning models to reach the final prediction—a method called model ensembles. In the pool of our selected papers, there were two papers leveraging this approach for metastasis detection. Abdollahi et al<sup>84</sup> and Munappa et al<sup>85</sup> ensemble models such as ResNet, DenseNet, and VGG on Camelyon datasets, Munnappa et al reported that ensemble model achieved higher precision, recall, and F1 score compared to single models. The speed of WSI processing was increased using reinforcement learning (RL) as suggested by Zhao et al.<sup>86</sup> The authors proposed the RLogist model, which tries to mimic the histopathologist's real-world approach in finding the regions of interest and analyzing those regions in multi-scale fashion. Gildenblat et al<sup>87</sup> investigated a self-supervised deep learning model, training a Siamese network. The authors leveraged the spatial continuity, treating adjacent tiles as more similar than distant ones, yielding a better tumor tile retrieval ratio compared to ordinary self-supervised models. A teacher–student model paradigm was investigated in a study by Cheng et al,<sup>88</sup> where the spatial distance was leveraged as tile similarity metric. Shubin et al<sup>89</sup> suggested including the variance error into deep learning model loss function, assuming that the variance error is included in the bias-variance trade-off. Challa et al<sup>90</sup> utilized a commercial software for LN metastasis detection offered by Visiopharm, achieving 100% sensitivity and relatively low specificity score of 41.5% in their breast cancer LN datasets.

Finally, there was a small percentage of studies ( $n = 3$ , 4%) describing non-deep learning AI methods for metastasis identification. Valkonen et al<sup>91</sup> compared patch classification task performance between a neural network and classic machine learning algorithms, such as support vector machine and random forest. Palatnik de Sousa et al<sup>92</sup> explored the explainability of the Camelyon16 dataset classification using locally interpretable model-agnostic explanations technique for Patch Camelyon patches split into superpixels by the SLIC algorithm. The follow-up study by the same authors<sup>93</sup> proposed genetic algorithm-inspired evolved explanation (EvEx) methodology to contrive model explanations and extract segmentation masks based on the explanation results.

#### Interpretation of LN metastasis misclassification

Only 12 studies (19%) out of 63 studies investigating LN metastasis provided a detailed analysis of false-positive predictions in metastatic LN detection. Based on the gathered data, LN sinus histiocytes were the most common false positively identified region mentioned by six studies<sup>29,31,56,68,90,94</sup> followed by secondary lymphoid follicles (germinal centers),<sup>56,64,94</sup> connective tissue,<sup>31,64</sup> out-of-focus areas,<sup>68</sup> and slide artifacts.<sup>50</sup> The most common metastatic regions misclassified as negative were micrometastatic lesions<sup>61,64</sup> as well as histiocyte-like tumor cells.<sup>53</sup>

Wang et al<sup>72</sup> identified adipose tissue, out-of-focus regions, and tumor edges as the areas where their model was struggling the most. To reduce false-positive rates for metastatic lesions, Pham et al<sup>44</sup> used a two-step deep learning pipeline including a follicle class to the metastasis detection model pipeline.

#### Segmentation of lymph node compartments

For this review, we found six (8%) articles focused on metastasis-free LN tissue characterization. Bekkhus et al<sup>16</sup> proposed a Mask-RCNN model for HEV segmentation in immunofluorescent stained LN samples called HEVfinder aiming at identification of dilated HEVs in tumor draining LNs in breast cancer. One study analyzed the clinical significance of extracellular matrix remodeling in metastatic LNs. Qaiser et al<sup>95</sup> studied tumor proximity to immunohistochemically detected collagen IV fibers in diffuse large B-cell lymphoma. The study utilized the HydraNet model to detect tumor cells and k-means clustering to identify the association strength between tumor cells and collagen regions in metastatic LNs. Kurian et al<sup>96</sup> proposed a U-Net derived models for LN sinus and germinal center segmentation implementing multi-scale and confidence map features into their deep learning solution. The authors implemented fuzzy boundaries to tackle imperfect sinus edge segmentations. A recent study by Jin et al<sup>97</sup> aimed

for detailed LN computational representation using 11 classes for LN segmentation. The study reused the patch-based classification approach using the LYNA model. This study aimed at real-time LN compartment predictions for augmented reality microscopy. The authors raised concerns about the choice of the optimal magnification for the model, illustrating this with tumor cell detection differences at 20x and 40x magnification. Verghese et al<sup>26</sup> trained multi-scale U-Net autoencoders with atrous convolution layers for LN sinus and germinal center detection in H&E-stained breast cancer slides. The authors trained the model using WSI crops at 2.5x, 5x and 10x magnifications, concluding that a mixture of different slide magnifications during training improved the final U-Net model performance. Song et al<sup>98</sup> utilized ResNet model to classify breast cancer LNs into either obese, metastatic, or metastasis-free achieving 0.67 AUC classification score. The slide tiles deemed the most representative by the model were further processed via rule base filtering to quantify adipocytes, erythrocytes, and lymphoid white space leading to a conclusion that increased size of all three components was observed in metastasis-free LNs of metastatic patients.

#### Model selection and experimental design

From 73 studies included in this review, 78% ( $n = 57$ ) were designed as patch classification experiments, whereas the remaining 22% ( $n = 16$ ) used pixel-level classification. In total, we found 27 distinct model architectures with 23 models describing patch classification and the remaining 4 models describing pixel-level segmentation studies. Among patch classifiers, ResNet models were utilized most frequently (15 studies),<sup>27,46–56,72,84,98</sup> followed by MIL ( $n = 9$ ),<sup>31,41,77–82,99</sup> VGG models ( $n = 8$ ),<sup>36,37,44,46,51,84,100,101</sup> and others as shown in Fig. 4B. Amongst pixel-level classification methods, U-Net was the most commonly used (10 studies)<sup>26,27,30,59–62,71,96,102</sup> followed by DeepLabv3 ( $n = 7$ ),<sup>28,29,62–66</sup> Mask-RCNN ( $n = 1$ ),<sup>16</sup> and proprietary Visiopharm AI metastasis model<sup>90</sup> as can be seen in Fig. 4B.

#### Chronological comparison of data and AI strategy selection

From a chronological perspective, patch-level classifier models pre-date pixel-level segmentation models, with the latter being increasingly used in recent years. Due to the popularity of the Camelyon challenges, earlier publications almost exclusively relied on Camelyon16, Camelyon17, and Patch Camelyon datasets, whereas publications with in-house datasets are increasingly emerging, see Fig. 4C. With respect to the research topic, studies focusing on metastasis-free LN compartment analysis emerged later than studies analyzing LN metastasis. All the studies on metastasis-free LNs use in-house dataset with the exception of Bekkhus et al, where a publicly available immunofluorescence dataset was used, and Gamez Serna et al, where rat mandibular LN dataset was utilized for LN contouring, see Table 1.

#### Discussion

There is a growing amount of evidence that AI methods are applicable for digital pathology. The launch of the Camelyon challenges has greatly accelerated the development of metastasis detection algorithms for LNs, with over 300 studies citing the original publication of the Camelyon17 dataset to this date. Consequentially, the publications offering deep learning solutions identified during this review of the literature were mostly focusing on LN metastasis detection. The majority of the published studies using the Camelyon datasets present novel deep learning model adaptations to improve metastasis detection without offering some biological interpretation of the model performance, in particular causes for false-positive or false-negative predictions. The models selected for metastasis identification in LNs included both patch classifiers and segmentation methods. ResNet models were most frequently used for patch classifiers, whereas U-Net models were most frequently selected for segmentation tasks. The majority of metastasis detection models were designed as patch classifiers (84%), suggesting that patch classifier output is sufficient to predict the LN status



(presence or absence of metastasis), and that refined tumor contour predictions obtained with segmentation methods may not be necessary for LN metastasis detection. From the selected papers, the best model performance on the Camelyon17 dataset was reported by Wang et al,<sup>66</sup> where a modified DeepLabv3+ model followed by DBSCAN clustering and XGBoost for pN stage prediction reached the weighted kappa score of 0.9632. However, the same study reported lower detection accuracy for ITCs. One of the options to reduce the difficulties of micrometastasis and ITC detection in LNs could be the use of AI to help the pathologist by highlighting the suspicious regions in the LN slide to the pathologist. Such a solution has been suggested by Google AI, developing the LYNA model designed specifically for pathological LN analysis.<sup>69</sup> The LYNA model has been designed to assist the pathologist by highlighting the potential areas with metastatic tumor allowing the pathologist to make the final decision for the analyzed slide.<sup>68</sup> In the same study, the LYNA-assisted workflow achieved a significantly higher sensitivity for micrometastasis detection compared to AI-unassisted H&E slide analysis by an experienced pathologist. To our knowledge, there's only one commercial software tool for LN metastasis detection provided by Visiopharm that is certified as an in vitro diagnostic medical device (CE IVD). In a study by Challa et al,<sup>90</sup> the use of Visiopharm metastasis algorithm resulted in a perfect tumor detection sensitivity of 100%, yet the reported precision was 41.5% suggesting a high rate of false-positive metastasis predictions.

Despite the relatively large number of studies focusing on metastasis detection in LNs included in our review (63 papers), there are currently relatively few studies utilizing AI models to answer clinically driven research questions offering a biological interpretation of the model results and/or aiming at model implementation into the clinic. 12 studies analyzed the metastasis detection model performance in the biological context to characterize the falsely detected metastatic regions. They concluded that histiocyte-filled LN sinuses and germinal centers are the most common false-positive metastasis predictions in LN slides. The deep learning models enabled researchers to systematically assess metastasis dissemination in the LNs. Wang et al<sup>27</sup> used a ResNet50 model to characterize the spatial patterns of metastatic spread in gastric cancer LNs, identifying two distinct mechanisms of metastasis based on the entry location of the tumor in the LN: via afferent lymphatic vessel or LN hilum. The authors trained a patch classifier at 20x magnification, which provided a sufficient output resolution to analyze the metastatic tumor location within LN achieving Dice score coefficient of 94.4%. To mitigate the putative false-positive predictions, the authors used LN sinus, germinal center, and adipose tissue patches as metastasis-free class during training. False-positive predictions in germinal centers could be tackled by developing a multi-class classification model that accounts for the germinal center class. This approach was implemented by Pham et al, who introduced a two-step deep learning pipeline combining two separate models for the detection of lymphoid follicles and metastasis regions.<sup>44</sup>

There is evidence accumulating to suggest that changes in secondary lymphoid follicles (germinal centers), sinuses, and paracortical areas of LN, so called reactive patterns (Fig. 1), may predict cancer patient survival. Allam et al<sup>45</sup> described a method for pN stage prediction from non-metastatic part of metastatic LNs called tumor microenvironment, indicating that interfollicular lymphocyte areas have the highest positive-predictive power for breast cancer LN cases. We previously summarized reactive pattern evaluation protocols suggested so far for manual assessment of reactive LN status, which highlighted the need for a standardized guideline for LN reaction assessment.<sup>103</sup> However, to establish evidence-based guidelines for LN reactive status evaluation, large scale biological studies need to be conducted to demonstrate its clinical value raising the necessity for automated detection methods. In this review, we identified six studies dedicated to AI-driven metastasis-free LN compartment analysis. The earliest study was published in 2019, highlighting the recency of this topic.<sup>95</sup> The LN sinuses tend to have a complex shape of canals penetrating LN cross-section in various directions, which can be harder to segment using a patch classification approach at lower magnification. This could be concluded from the study by Jin et al,<sup>97</sup> where the authors trained

LYNA models for LN compartment detection at different magnifications. The highest accuracy for LN sinus segmentation was achieved analyzing patches at 40x magnification, whereas 10x magnification was optimal for breast cancer metastasis patch identification. In the same study, veins, arteries, nerves, LN capsule, fat, and lymphocytes were also best detected at 40x magnification except for germinal centers, where best test performance was achieved at 20x magnification. Manual LN sinus annotation proved to be a challenging task as suggested by Kurian et al, where authors implemented a U-Net autoencoder model for LN sinus segmentation accounting for noisy LN sinus annotations by developing a loss function for fuzzy boundaries.<sup>96</sup> In a follow-up study from the same research group, the authors also explored the post-processing of the deep learning model output and identified evaluation metrics that could be translated into the clinical practice. The germinal center predictions were used to obtain metrics such as germinal center count per LN, germinal center mean area and germinal center circularity using the Polsby-Popper circularity metric for shape evaluation. For LN sinuses, the subcapsular sinus width per LN was evaluated. The study revealed that a larger number of germinal centers, a larger germinal center area, and its rounder shape, as well as wider subcapsular LN sinuses were associated with longer survival. However, these results require validation in studies in LNs from other types of cancer. In a recent study by Song et al,<sup>98</sup> authors indicated the importance of prognostic LN structure size evaluation, concluding that the size of adipocytes, red blood cell conglomerates, and white gaps in LN have a prognostic significance when distinguishing patients by pN status and obesity stage. There was only a single study where the segmentation of HEVs was attempted. Bekkhus et al<sup>16</sup> utilized a deep learning model to segment HEVs in immunofluorescent LN images. Besides HEV detection, the authors also aimed at HEV dilatation detection, as it was previously suggested that this might be associated with pre-metastatic changes in regional LNs.<sup>104</sup> Further analysis is needed to validate whether HEV detection and dilation evaluation can be assessed in routine H&E-stained slides.

When designing an AI study, the selection of the deep learning model plays a fundamental role. The majority of the studies included in this review describe an adaptation of already existing model architectures for the LN analysis task and the most commonly used models are ResNet, DenseNet, and U-Net. The LYNA model introduced by Google AI was specifically designed to tackle the metastatic LN identification task, which was later adapted for metastasis-free LN compartment segmentation.<sup>69,70,97</sup> The authors of LYNA model applied the multi-scale model paradigm, where the model accepts multiple input images at different magnifications, encompassing both low- and high magnification level features, to the LN analysis field. The multi-scale LN model idea was further developed by Schmitz et al,<sup>71</sup> proposing a combination of an U-Net autoencoder and ResNet18 context encoder models. Gamez Serna et al<sup>24</sup> introduced multi-scale MMO-Net model capable of multi-organ segmentation, including LNs. In a recent study by D'Amato et al,<sup>105</sup> multi-scale models were proven to achieve higher F1 score metrics for tumor detection in 20 different cancer types compared to single-magnification setting, however the input-output intensive operations render multi-scale models timewise inefficient at WSI processing. The time efficiency issue has been addressed by Wang et al,<sup>72</sup> proposing a model that reduced slide processing time more than eight times compared to LYNA model.

Dataset selection is another key factor when designing the computational study. Except for MMO-Net dataset released for rat mandibular LN contour detection, all the public datasets included in our review were exclusively generated from breast cancer LN cases with the ground truth specifically oriented towards metastasis evaluation (except HEV immunofluorescence dataset), which limits the scope of their usability for other research objectives. There is a need for alternative data sources. The low incidence of external validation of models found in the selected study pool (29%,  $N=21$ ) is most likely related to the apparent lack of publicly available data that could be used as a ground truth in external validation. The Camelyon17 testing dataset contains slides from multiple different medical centers, addressing the practical challenge of site-specific WSI features, however, the slides originate from a single country

(Netherlands), thus limiting the capabilities to evaluate model generalizability over different patient geographies.

The datasets used in the studies had a varying degree of public accessibility. Several datasets, such as OE02 esophageal cancer trial data as well as breast cancer LN dataset from Guy's hospital had only images publicly available with inaccessible annotations, whereas only the datasets with accessible ground truth were regarded as publicly available in our review. The datasets varied in the nature of ground truth format—four datasets contained manually annotated ground truth, whereas AIDA breast LN dataset was complemented with immunohistochemically stained slides for metastasis. Compared to the manual annotations, immunohistochemically stained ground truth offers the highest annotation recall, as this method avoids the bias of human-produced errors. However, artifacts such as IHC stain residues might obfuscate the slide translation into ground truth mask, raising the need for pathologist to review the stained slides.<sup>94</sup> Additionally, not all the anatomical structures of interest have a specific antibody that would produce annotation-quality staining. For instance, LN sinus macrophages have a specific stain for CD169+ macrophage cells, however, only a small subset of sinus-residing macrophages express CD169 molecule, whereas CD68 marker for macrophages is not specific for this cell type only,<sup>106</sup> which would produce false positive staining. Due to these issues, the ground truth annotation process requires an expert-in-a-loop to ensure reliable outcome. However, the laborious annotation process has a potential to be at least partially automated by applying novel annotation tools that implement active learning, such as Monai Label, DeePathology, or DigiPath Viewer.

#### Limitations

With the exception of one study,<sup>16</sup> studies included in this review were limited to those using light microscopy imaging data because the primary focus of this review was the analysis of routine H&E-stained slides. However, numerous studies have been performed on other sample analysis methods. Miura et al<sup>107</sup> have employed scanning acoustic microscopy to identify lesion elasticity in LNs. Liu et al<sup>108</sup> relied on multi-spectral microscopy to obtain elliptic features of cells, enabling LN segmentation into tumor, B, T, and dendritic cells. Fourier transform infrared (FTIR) microscopy has been applied for LN architecture analysis, leading to complete unsupervised LN architecture reconstruction by Bird et al,<sup>109</sup> followed by later study by Leslie et al<sup>110</sup> rendering higher-resolution LN microarchitecture features of eight distinct classes. Compared to light microscopy, each analysis method has specific advantages: scanning acoustic microscopy is capable of analyzing tissue elasticity, multi-spectral microscopy includes spectral bands in infrared and ultraviolet wavelengths besides red, green, and blue color channels, FTIR method captures cell-specific molecular composition, thus offering higher resolution than H&E stained tissue analysis. These methods, despite their own limitations, could complement the traditional H&E sample analysis by adding new insights into tissue composition and biochemical processes inside LNs.

This review was limited to the studies analyzing LNs without further investigating the structural composition of the primary tumor. Tertiary lymphoid structures (TLSs) found inside and nearby the primary tumor are gaining scientific attention. In the study by Ling et al,<sup>111</sup> Inception-ResNet-v2 model was trained for patch classification to detect TLSs in esophageal cancer patients and to distinguish between immature and mature TLS cases. In the same study, mature TLSs have been associated with a better patient disease-free survival. Barmpoutis et al<sup>112</sup> have published their approach for TLS segmentation in lung cancer patients using DeepLabv3+ and Inception-ResNet-v2 for feature extraction.

The heterogeneity of datasets and their usage in the studies have limited our capabilities to perform a meta-analysis on the included studies, limiting the quantitative power of study comparison. For the metastasis detection task, 68% ( $n=43$ ) of all included metastasis detection studies used one or more of five publicly available datasets, whereas 39% ( $n=25$ ) of studies employed inhouse, publicly inaccessible datasets with or without mixing it with publicly available data. In case of metastasis-free LN analysis, five

out of six reviewed studies used different publicly unavailable datasets. We considered that quantitative AI method comparison for models trained and tested on different datasets would carry a limited value, thus we refrained from quantitative data interpretation across the reviewed articles. Another limitation of this study was the language, as only the articles published in English were included, thus constricting the range of articles screened for our analysis.

#### Conclusions

Based on the published evidence gathered for this review, it can be concluded that the computational lymph node analysis field is currently evolving. The LN metastasis detection task has been extensively explored by the scientific community, leading to promising commercial solutions such as Visiopharm metastasis app, Paige Breast Lymph Node system. However, when considering clinical implementation, challenges such as sensitivity and specificity maximization, as well as model robustness evaluation via external testing sets need to be addressed. These practical challenges highlight the current limitations of accessible annotated LN datasets. New research studies are emerging acknowledging the computational characterization of metastasis-free LN compartment rearrangements, resulting in first attempts to segment LN compartments in H&E-stained slides including LN sinuses, germinal centers, vessels, and paracortical lymphocyte areas. Further research is warranted to establish and validate the clinically interpretable metrics describing the features of segmented LN structures.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

H.W. has minority shares in the company Radiomics SA. D.R.M. is a director of HeteroGenius Limited. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

HIG and DRM are supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpi.2024.100367>.

#### References

- Nathanson SD, Kwon D, Kapke A, Alford SH, Chitale D. The role of lymph node metastasis in the systemic dissemination of breast cancer. *Ann Surg Oncol* 2009;16(12):3396–3405.
- Kinami S, Saito H, Takamura H. Significance of lymph node metastasis in the treatment of gastric cancer and current challenges in determining the extent of metastasis. *Front Oncol* 2022;11, 806162.
- Matsuda S, Takeuchi M, Kawakubo H, Kitagawa Y. Lymph node metastatic patterns and the development of multidisciplinary treatment for esophageal cancer. *Dis Esophagus* 2023;36(4).
- Kim HJ, Choi GS. Clinical implications of lymph node metastasis in colorectal cancer: current status and future perspectives. *Ann Coloproctol* 2019;35(3):109–117.
- Gusterson BA. The new TNM classification and micrometastases. *Breast* 2003;12(6):387–390.
- Riegrova D, Jansa P. Prognostic significance of reactive changes in regional lymph nodes in gastric and mammary carcinomas. *Neoplasma* 1982;29(4):481–486.
- Hartveit F. Reactive changes in lymph-nodes draining breast-carcinoma and their relation to tumor spread. *Int J Oncol* 1993;2(6):969–976.
- Saldanha P. Morphological assessment of lymph nodes draining carcinoma. *MGM J Med Sci* 2016;3:190–197.
- Murray D, Hreno A, Dutton J, Hampson LG. Prognosis in colon cancer: a pathologic reassessment. *Arch Surg* 1975;110(8):908–913.

10. Friedell GH, Soto EA, Kumaoka S, Abe O, Hayward JL, Bulbrook RD. Sinus histiocytosis in British and Japanese patients with breast cancer. *Lancet* 1974;304(7891):1228–1229.
11. Bedikian AY, Chen TT, Khankhanian N, et al. The natural history of gastric cancer and prognostic factors influencing survival. *J Clin Oncol Off J Am Soc Clin Oncol* 1984;2(4):305–310.
12. Black MM, Barclay THC, Hankey BF. Prognosis in breast cancer utilizing histologic characteristics of the primary tumor. *Cancer* 1975;36(6):2048–2055.
13. Black MM, Freeman C, Mork T, Harvei S, Cutler SJ. Prognostic significance of microscopic structure of gastric carcinomas and their regional lymph nodes. *Cancer* 1971;27(3):703–711.
14. Zhou X-G, Yu B-M, Shen Y-X. Surgical treatment and late results in 1226 cases of colorectal cancer. *Dis Colon Rectum* 1983;26(4):250–256.
15. Kodama Y, Seojima K, Inokuchi K. Reactive hyperplasia of paracortex and germinal centers of the regional lymph nodes in gastric carcinoma cases as a favorable prognostic indicator. *Jpn J Surg* 1976;6(1):9–18.
16. Bekkhus T, Martikainen T, Olofsson A, et al. Remodeling of the lymph node high endothelial venules reflects tumor invasiveness in breast cancer and is associated with dysregulation of perivascular stromal cells. *Cancers [Internet]* 2021;13(2).
17. Litjens G, Bandi P, Ehteshami Bejnordi B, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience* 2018;7(6).
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 2021;10(1):89.
19. Mongan J, Moy L, Charles E, Kahn J. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intel* 2020;2(2), e200029.
20. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318(22):2199–2210.
21. Veeling B, Linmans J, Winkens J, Cohen T, Welling M. *Rotation Equivariant CNNs for Digital Pathology: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II*. 09, 2018:210–218.
22. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;25(8):1301–1309.
23. Sofia Jarkman ML, Hedlund J, Treanor D, Lundstrom C, van der Laak J. *Axillary Lymph Nodes in Breast Cancer Cases*. 2019.
24. Gámez Serna C, Romero-Palomo F, Arcadu F, Funk J, Schumacher V, Janowczyk A. MMO-Net (Multi-Magnification Organ Network): a use case for organ identification using multiple magnifications in preclinical pathology studies. *J Pathol Inform* 2022;13.
25. Niemiisto A, Shmulevich I, Yli-Harja O, Chiriac LR, Hamilton SR. Automated quantification of lymph node size and number in surgical specimens of stage II colorectal cancer. *Conf Proc IEEE Eng Med Biol Soc* 2005;2005:6313–6316.
26. Verghese G, Li M, Liu F, et al. Multiscale deep learning framework captures systemic immune features in lymph nodes predictive of triple negative breast cancer outcome in large-scale studies. *J Pathol* 2023;260(4):376–389.
27. Wang X, Chen Y, Gao Y, et al. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. *Nat Commun* 2021;12(1):1637.
28. Huang SC, Chen CC, Lan J, et al. Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. *Nat Commun* 2022;13(1):3347.
29. Hu Y, Su F, Dong K, et al. Deep learning system for lymph node quantification and metastatic cancer identification from whole-slide pathology images. *Gastric Cancer* 2021;24(4):868–877.
30. Beuque M, Magee DR, Chatterjee A, et al. Automated detection and delineation of lymph nodes in haematoxylin & eosin stained digitised slides. *J Pathol Inform* 2023;14, 100192.
31. Tan L, Li H, Yu J, et al. Colorectal cancer lymph node metastasis prediction with weakly supervised transformer-based multi-instance learning. *Med Biol Eng Comput* 2023;61(6):1565–1580.
32. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8(1):53.
33. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al, eds. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015.
34. He K, Zhang X, Ren S, Sun J, eds. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016.
35. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, eds. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2017.
36. Wang D, Khosla A, Gargeya R, Irshad H, Beck A. *Deep Learning for Identifying Metastatic Breast Cancer*. 2016.
37. Jamaluddin MF, Fauzi MFA, Abas FS. *Tumor detection and whole slide classification of H&E lymph node images using convolutional neural network*. 2017:90–95.
38. Lin H, Chen H, Dou Q, Wang L, Qin J, Heng P-A. ScanNet: A Fast and Dense Scanning Framework for Metastatic Breast Cancer Detection from Whole-Slide Images. 032018.
39. Lin H, Chen H, Graham S, Dou Q, Rajpoot N, Heng PA. Fast ScanNet: fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection. *IEEE Trans Med Imaging* 2019;38(8):1948–1958.
40. Zhang W, Zhu C, Liu J, Wang Y, Jin M. Cancer metastasis detection through multiple spatial context network. *Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition*. (New York, NY, USA): Association for Computing Machinery; 2020. p. 221–225.numpages = 5.
41. Courtiol P, Tramel E, Samselme M, Wainrib G. *Classification and Disease Localization in Histopathology Using Only Global Labels: A Weakly-Supervised Approach*. 2018.
42. Shvetsova N, Bakker B, Fedulova I, Schulz H, Dylvov DV. Anomaly detection in medical imaging with deep perceptual autoencoders. *IEEE Access* 2021;9:118571–118583.
43. Alheejawi S, Xu H, Berendt R, Jha N, Mandal M. Novel lymph node segmentation and proliferation index measurement for skin melanoma biopsy images. *Comput Med Imaging Graph* 2019;73:19–29.
44. Pham HHN, Futakuchi M, Bychkov A, Furukawa T, Kuroda K, Fukuoka J. Detection of lung cancer lymph node metastases from whole-slide histopathologic images using a two-step deep learning approach. *Am J Pathol* 2019;189(12):2428–2439.
45. Allam K, Wang XI, Zhang S, et al. Deep Learning Provides Rapid Screen for Breast Cancer Metastasis with Sentinel Lymph Nodes. *arXiv preprint arXiv:230105938*. 2023.
46. Jaiswal A, Panshin I, Shulkin D, Aneja N, Abramov S. *Semi-Supervised Learning for Cancer Detection of Lymph Node Metastases*. 2019.
47. Kim YG, Kim S, Cho CE, et al. Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections. *Sci Rep* 2020;10(1):21899.
48. Chen C, Zheng S, Guo L, et al. Identification of misdiagnosis by deep neural networks on a histopathologic review of breast cancer lymph node metastases. *Sci Rep* 2022;12(1):13482.
49. A robust and effective approach towards accurate metastasis detection and pn-stage classification in breast cancer. In: *Lee B, Paeng K, eds. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II*. 11. Springer; 2018.
50. Patil A, Diwakar H, Sawant J, et al. Efficient quality control of whole slide pathology images with human-in-the-loop training. *J Pathol Inform* 2023;14, 100306.
51. Turki T, Al-Sharif A, Taguchi Yh. *End-to-End Deep Learning for Detecting Metastatic Breast Cancer in Axillary Lymph Node from Digital Pathology Images*, 11. 2021:343–353.
52. Chuang WY, Chen CC, Yu WH, et al. Identification of nodal micrometastasis in colorectal cancer using deep learning on annotation-free whole-slide images. *Mod Pathol* 2021;34(10):1901–1911.
53. Matsushima J, Sato T, Yoshimura Y, et al. Clinical utility of artificial intelligence assistance in histopathologic review of lymph node metastasis for gastric adenocarcinoma. *Int J Clin Oncol* 2023;28(8):1033–1042.
54. Kronberger RM, Haerberle L, Pfaus M, et al. Communicator-driven data preprocessing improves deep transfer learning of histopathological prediction of pancreatic ductal adenocarcinoma. *Cancers (Basel)* 2022;14(8).
55. Matsushima J, Sato T, Ohnishi T, et al. The use of deep learning-based computer diagnostic algorithm for detection of lymph node metastases of gastric adenocarcinoma. *Int J Surg Pathol* 2023;31(6):975–981.10.668969221113475.
56. Tang H, Li G, Liu C, et al. Diagnosis of lymph node metastasis in head and neck squamous cell carcinoma using deep learning. *Laryngoscope Investig Otolaryngol* 2022;7(1):161–169.
57. Ronneberger O, Fischer P, Brox T, eds. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*. 18. Springer; 2015.
58. Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:170605587*; 2017.
59. Lin Xu CX, Y Tong, YC Su. Detection and Classification of Breast Cancer Metastases Based on U-Net2019.
60. Jin YW, Jia S, Ashraf AB, Hu P. Integrative data augmentation with U-net segmentation masks improves detection of lymph node metastases in breast cancer patients. *Cancers (Basel)* 2020;12(10).
61. Jansen P, Bager DO, Duschner N, et al. Deep learning detection of melanoma metastases in lymph nodes. *Eur J Cancer* 2023;188:161–170.
62. Olga A, Mainovskaya SIA, Devyatkin AV, et al. Development and application of artificial intelligence for the detection of the lymph nodes metastases in colorectal cancer. *Clinical Guidelines* 2022:49–59.
63. Khalil MA, Lee YC, Lien HC, Jeng YM, Wang CW. Fast segmentation of metastatic foci in H&E whole-slide images for breast cancer diagnosis. *Diagnostics (Basel)* 2022;12(4).
64. Pan Y, Sun Z, Wang W, et al. Automatic detection of squamous cell carcinoma metastasis in esophageal lymph nodes using semantic segmentation. *Clin Transl Med* 2020;10(3), e129.
65. Bozdağ Z, Talu F. Pyramid nonlocal network for histopathological image of breast lymph node segmentation. *Int J Computat Intel Syst* 2020:14.
66. Wang L, Song T, Katayama T, Jiang X, Shimamoto T, Leu J-S. Deep regional metastases segmentation for patient-level lymph node status classification. *IEEE Access* 2021;9:129293–129302.
67. A deep learning approach for semantic segmentation in histology tissue images. In: *Wang J, JD MacKenzie, Ramachandran R, Chen DZ, eds. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II*. 19. Springer; 2016.
68. Liu Y, Kohlberger T, Norouzi M, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Arch Pathol Lab Med* 2019;143(7):859–868.
69. Liu Y, Gadepalli K, Norouzi M, et al. *Detecting Cancer Metastases on Gigapixel Pathology Images*. 2017.
70. Steiner DF, MacDonald R, Liu Y, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am J Surg Pathol* 2018;42(12):1636–1646.
71. Schmitz R, Madesta F, Nielsen M, et al. Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. *Med Image Anal* 2021;70, 101996.
72. Wang R, Gu Y, Zhang T, Yang J. Fast cancer metastasis location based on dual magnification hard example mining network in whole-slide images. *Comput Biol Med* 2023;158, 106880.
73. Kovalev V, Kazlouki S. *Examining the Capability of GANs to Replace Real Biomedical Images in Classification Models Training*. 2019.

74. Xue Y, Ye J, Zhou Q, et al. Selective synthetic augmentation with HistoGAN for improved histopathology image classification. *Med Image Anal* 2021;67, 101816.
75. Wollmann T, Eijkman CS, Rohr K. *Adversarial domain adaptation to improve automatic breast cancer grading in lymph nodes*. 2018:582–585.
76. Dimitriou N, Arandjelović O, Caie PD. Deep learning for whole slide image analysis: an overview. *Front Med* 2019;6:264.
77. Akbar S, Martel A. *Cluster-Based Learning from Weakly Labeled Bags in Digital Pathology*. 2018.
78. Shao Z, Bian H, Chen Y, et al. *TransMIL: Transformer Based Correlated Multiple Instance Learning for Whole Slide Image Classification*. 2021.
79. Wang Q, Zou Y, Zhang J, Liu B. Second-order multi-instance learning model for whole slide image classification. *Phys Med Biol* 2021;66(14).
80. Kang H, Yang M, Zhang F, et al. Identification lymph node metastasis in esophageal squamous cell carcinoma using whole slide images and a hybrid network of multiple instance and transfer learning. *Biomed Signal Process Control* 2023;82, 104577.
81. Sadafi A, Navab N, Marr C. Active learning enhances classification of histopathology whole slide images with attention-based multiple instance learning. *arXiv preprint arXiv:230301342*; 2023.
82. Yu J-G, Wu Z, Ming Y, et al. Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images. *Med Image Anal* 2023;85, 102748.
83. Qin W, Xu R, Jiang S, Jiang T, Luo L, eds. PathTR: Context-Aware Memory Transformer for Tumor Localization in Gigapixel Pathology Images. *Proceedings of the Asian Conference on Computer Vision*; 2022.
84. Jafar Abdullahi ND, Panahi Y, Gardaneh M. Detection of metastatic breast cancer from whole-slide pathology images using an ensemble deep-learning method. *Arch Breast Cancer* 2022;9(3):364–376.
85. Munappa S, Subhashini J, Pallikonda SS. Minimizing false negatives in metastasis prediction for breast cancer patients through a deep stacked ensemble analysis of whole slide images. *Traitement du Signal* 2023;40(3):1289.
86. Zhao B, Zhang J, Ye D, et al. RLogist: fast observation strategy on whole-slide images with deep reinforcement learning. *arXiv preprint arXiv:221201737*; 2022.
87. Gildenblat J, Klaiman E. *Self-Supervised Similarity Learning for Digital Pathology*. 2019.
88. Cheng H-T, Yeh C-F, Kuo P-C, et al. *Self-similarity Student for Partial Label Histopathology Image Segmentation*. 2020.
89. Shubin D, Eytan D, Goodfellow S. *About Explicit Variance Minimization: Training Neural Networks for Medical Imaging With Limited Data Annotations*. 2021.
90. Challa B, Tahir M, Hu Y, et al. Artificial intelligence-aided diagnosis of breast cancer lymph node metastasis on histologic slides in a digital workflow. *Mod Pathol* 2023;36(8), 100216.
91. Valkonen M, Kartasalo K, Liimatainen K, Nykter M, Latonen L, Ruusuvoori P. Metastasis detection from whole slide images using local features and random forests. *Cytometry A* 2017;91(6):555–565.
92. Palatnik de Sousa I, Maria Bernardes Rebuzzi Vellasco M, Costa da Silva E. Local interpretable model-agnostic explanations for classification of lymph node metastases. *Sensors (Basel)* 2019;19(13).
93. Palatnik de Sousa I, Vellasco M, Costa da Silva E. Evolved explainable classifications for lymph node metastases. *Neural Netw* 2022;148:1–12.
94. Jarkman S, Karlberg M, Pocevičiūtė M, et al. Generalization of deep learning in digital pathology: experience in breast cancer metastasis detection. *Cancers* 2022;14(21):5424.
95. Qaiser T, Pugh M, Margielewska-Davies S, Hollows R, Murray P, Rajpoot N. *Digital Tumor-Collagen Proximity Signature Predicts Survival in Diffuse Large B-Cell Lymphoma*. 2019:163–171.
96. Kurian N, Lohan A, Verghese G, et al. *Deep Multi-Scale U-Net Architecture and Noise-Robust Training Strategies for Histopathological Image Segmentation*. 2022.
97. David J, Joseph HR, Elaine ET, et al. Independent assessment of a deep learning system for lymph node metastasis detection on the Augmented Reality Microscope. *J Pathol Inform* 2022;13, 100142.
98. Song Q, Muller KE, Hondelink LM, Karagas M, Hassanpour S. Non-metastatic axillary lymph nodes have distinct morphology and immunophenotype in obese breast cancer patients at risk for metastasis. *Am J Pathol* 2024;194(2):253–263.
99. Wang Z, Saoud C, Wangsiricharoen S, James AW, Popel AS, Sulam J. Label cleaning multiple instance learning: refining coarse annotations on single whole-slide images. *IEEE Trans Med Imaging* 2022;41(12):3952–3968.
100. Kim YG, Song IH, Lee H, et al. Challenge for diagnostic assessment of deep learning algorithm for metastases classification in sentinel lymph nodes on frozen tissue section digital slides in women with breast cancer. *Cancer Res Treat* 2020;52(4):1103–1111.
101. Luz DS, Costa RJO, Rabêlo RdAL, Rodrigues JJPC, Araujo FHD. *Automatic Identification of Metastasis in Histopathological Images Using Deep Learning*. 2021:1–6.
102. Khan A, Janowczyk A, Müller F, et al. Impact of scanner variability on lymph node segmentation in computational pathology. *J Pathol Inform* 2022;13, 100127.
103. Budginaite E, Kloft M, van Kuijk SMJ, et al. The clinical importance of the host anti-tumour reaction patterns in regional tumour draining lymph nodes in patients with locally advanced resectable gastric cancer: a systematic review and meta-analysis. *Gastric Cancer* 2023;26(6):847–862.
104. Shen H, Wang X, Shao Z, et al. Alterations of high endothelial venules in primary and metastatic tumors are correlated with lymph node metastasis of oral and pharyngeal carcinoma. *Cancer Biol Ther* 2014;15(3):342–349.
105. D'Amato M, Szostak P, Torben-Nielsen B. A comparison between single-and multi-scale approaches for classification of histopathology images. *Front Public Health* 2022;10, 892658.
106. Ohnishi K, Komohara Y, Saito Y, et al. CD169-positive macrophages in regional lymph nodes are associated with a favorable prognosis in patients with colorectal carcinoma. *Cancer Sci* 2013;104(9):1237–1244.
107. Miura K, Nasu H, Yamamoto S. Scanning acoustic microscopy for characterization of neoplastic and inflammatory lesions of lymph nodes. *Scient Rep* 2013;3:1255.
108. Liu X, Setiadi A, Alber M, Lee P, Chen D. Identification and classification of cells in multi-spectral microscopy images of lymph nodes. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, 7962 ; 2011.
109. Bird B, Mijlkovic M, Romeo MJ, et al. Infrared micro-spectral imaging: distinction of tissue types in axillary lymph node histology. *BMC Clin Pathol* 2008;8:8.
110. Leslie LS, Wrobel TP, Mayerich D, Bindra S, Emmadi R, Bhargava R. High definition infrared spectroscopic imaging for lymph node histopathology. *PLoS One* 2015;10(6), e0127238.
111. Ling Y, Zhong J, Weng Z, et al. The prognostic value and molecular properties of tertiary lymphoid structures in oesophageal squamous cell carcinoma. *Clin Transl Med* 2022;12(10), e1074.
112. Barmpoutis P, Di Capite M, Kayhanian H, et al. Tertiary lymphoid structures (TLS) identification and density assessment on H&E-stained digital slides of lung cancer. *PLoS One* 2021;16(9), e0256907.